

An Intelligent Computerized Search Engine

BACKGROUND OF THE INVENTION

1. Field of Invention

This present invention relates to query processing, and more specifically relates to techniques for identifying entries that are conceptually similar to the search criteria.

2. Description of Related Art

With the increasing popularity of the Internet and the World Wide Web, a large number of highly specialized sites have come on line that exclusively address very narrowly defined subject matter. Their applications range from obscure technical disciplines to specialty e-commerce merchants. Most, however, maintain their information in databases that contain descriptive phrases in each record. This architecture allows the sites to provide search engines intended to help on-line users easily locate their desired information.

The vast majority of current search engines are fundamentally based on a direct character string comparison function. When a user submits a query containing one or more query terms, the search engine identifies records that contain character strings that are exact matches to the query terms. While many current search engines supplement this basic functionality with Boolean capabilities and "wildcard" characters, the search itself is precisely literal. An exhaustive set of matching citations is returned for user review. In the hands of a sophisticated user, fluent in the exact terminology of the database, these search engines can efficiently highlight the desired information. Small variations in nomenclature, however, are catastrophic

OCT-16-00 MON 07:06 AM BECKER CAPITAL

FAX NO. 3035272799

P. 05/26

for the underlying matching function. For example, a user seeking information on "bikes" will not be shown references to "bicycles". As a result, novice users often miss many relevant records due to the limitations of the underlying character string matching function.

An alternative approach to this situation is to force the descriptions and query terms into a standardized set of categories (fields) and entries (allowed terms). The resulting structured query is often executed using "drop down" boxes that limit input to acceptable inputs. This rigid approach has discouraged its use by many novices and still fails to identify matches when the terminology of the database is not intuitively obvious to the casual observer.

In an attempt to allow more natural unstructured user input, a number of search engines have been developed that attempt to search based on the contents, or semantics, of the query. The direct application of this approach has not been successful due to the ambiguous and contextually specific nature of natural language (i.e. "cycling" may refer to riding a bicycle, riding a motorcycle or repeating the same set of actions, depending on the context). Further, these engines remain completely intolerant of the kind of partially incorrect input that is typical of novice users. The proliferation of highly specialized databases, however, offers the opportunity to exploit their coverage of only a very limited domain of information. This allows a minimal vocabulary and a single predominate semantic structure to effectively characterize the content of the domain.

Consequently, the prior art does not provide the novice with a means to intuitively search specialized databases with just a layman's vocabulary and only a partial understanding of the subject matter. This failure has substantial commercial significance for a number of Internet businesses, such as electronic auctions. These businesses cater to a wide variety of consumers that typically include many "novice" users. Given the fiercely competitive nature of the industry, even minor inconveniences in the user interface will move customers from one web business to another. ("Your competition is only a click away") Once a consumer has chosen a web auction, potential buyers and sellers of a particular item must find each other to initiate a negotiation. Given the breadth of items offered at any one time, search engines are typically employed by potential buyers to identify offers of interest. The limitations of existing search engines cause them to miss potential matches and preclude potential sales.

SUMMARY OF THE INVENTION

To provide a means for a novice user to quickly and easily identify records of interest in a specialized database, without specific knowledge of the covered subject matter.